Technical Report TR-67-44 March 1967
NsG-398

A Chebyshev Method for the Numerical Solution of the One-dimensional Heat Equation

by

J. C. Mason

Computer Science Center
University of Maryland

Abstract

A generalization of the Lanczos tau method is described for the numerical solution of a class of heat conduction problems. It is shown to be more efficient than the Crank-Nicolson method for some typical examples.

Introduction

In a previous paper (Mason, 1967) we have described a two-dimensional Chebyshev method for the solution of partial differential equations over bounded regions for certain types of boundary problems. In this paper we shall consider the one-dimensional heat equation and derive a Chebyshev method which is essentially one-dimensional.

1. The Heat Conduction Problem

Consider the heat conduction equation of the form

$$v_{xx} - v_t = 0 \qquad (-1 < x < 1,\ o \leq t < \infty) \qquad \ldots(1)$$

with boundary conditions

$$\left.\begin{array}{lr} \lambda_{-1}\, v + \mu_{-1}\, v_x = \phi_{-1}\,(t) & \text{on } x = -1, \\ \lambda_1\, v + \mu_1\, v_x = \phi_1\,(t) & \text{on } x = 1, \\ v = f(x) & \text{on } t = o. \end{array}\right\} \quad \ldots(2)$$

In the usual notation v_x and v_t denote the partial derivatives of v with respect to x and t respectively. The functions ϕ_{-1} (t), ϕ_1 (t), f (x) and the constants λ_{-1}, μ_{-1}, λ_1, μ_1, are given, and the diffusivity of the substance has been taken as unity by a suitable transformation of t.

The two methods which are usually advocated for solving such problems are the Fourier series method and the implicit finite difference method of Crank and Nicolson (1947). The Crank-Nicolson scheme for the problem (1), (2) is described concisely in Smith (1965). The Fourier method, which is described in detail in Carslaw and Jaeger (1959), leads to infinite analytical expansions for the solution $v(x, t)$. Some typical examples of such expansions are as follows.

<u>Example (i)</u>

$$\lambda_{-1} = \lambda_1 = 1; \; \mu_{-1} = \mu_1 = 0; \; \phi_{-1} \equiv \phi_1 \equiv 0.$$

Suppose $f(x)$ is an odd function of x with a Fourier expansion

$$f(x) = \sum_{k=1}^{\infty} a_k \sin k\pi x,$$

where $a_k = \int_{-1}^{1} f(x) \sin k\pi x \, dx.$

Then $v(x, t) = \sum_{k=1}^{\infty} a_k e^{-k^2\pi^2 t} \sin k\pi x.$

<u>Example (ii)</u>

$$\lambda_{-1} = \lambda_1 = h; \; \mu_{-1} = -1, \; \mu_1 = 1; \; \phi_{-1} \equiv \phi_1 \equiv 0.$$

Suppose α_k is the k^{th} positive zero of

$$(\alpha^2 - h^2) \tan 2\alpha = 2\alpha h,$$

and that $c_k = h \sin \alpha_k + \alpha_k \cos \alpha_k,$

$$d_k = h \cos \alpha_k - \alpha_k \sin \alpha_k,$$

$$g_k(x) = c_k \cos \alpha_k x + d_k \sin \alpha_k x.$$

Then $v(x,t) = \sum_{k=1}^{\infty} (\alpha_k + h^2 + h)^{-1} a_k e^{-\alpha_k^2 t} g_k(x),$ ...(3)

where $a_k = \int_{-1}^{1} f(x) g_k(x) \, dx.$ ...(4)

This expansion is valid when $f(x)$ has an infinite expansion in the functions $\{g_k(x)\}$, which reduces to a "traditional" Fourier series expansion in the cases $h = 0$ and $h = \infty$.

The presence of non-trivial functions $\phi_{-1}(t)$ and $\phi_1(t)$ in the boundary conditions (2) introduces terms like

$$e^{-\beta_k^2 t} \int_0^t e^{\beta_k^2 s} \left\{ \phi_{-1}(s) - (-1)^k \phi_1(s) \right\} ds \qquad \ldots(5)$$

into the expansions.

For numerical computations the relevant expansion would be terminated after a certain number, say n, of terms. The resulting approximation could then be made explicit by calculating n definite integrals like (4) and n indefinite integrals like (5). But, in general, separate expansions are needed for each distinct line in the t-direction, since the integrals (5) require numerical quadrature. Finite difference methods like the Crank-Nicolson method are specifically designed to make use of steps in the t-direction, and the simplicity of such methods appears to give them a clear advantage over the Fourier series method. Elliott (1961) has, however, streamlined the Fourier method into a Chebyshev method which could offer stronger competition. Elliott uses finite difference approximations in the t-direction to obtain approximations of form $\sum a_i T_i(x)$ to $v(x,t)$ for each t. Using the standard notation, $T_i(x)$ denotes the Chebyshev polynomial of degree i, which is appropriate to the range $[-1, 1]$ of x.

This is not the end of the matter, however, since there are many problems for which the integrals (5) can be evaluated.

explicitly from exact or approximate analytical expressions. In particular this happens when $\emptyset_{-1}$ (t) and $\emptyset_1$ (t) can be adequately represented by sums of products of functions like t, e^{-t}, cos t, and sin t. In such cases we believe that the Fourier method is generally more efficient than Crank-Nicolson, because it does not involve any step-by-step integration in the t-direction. Moreover, advantage has been taken of the exponential behavior of v (x,t) in the adopted form of approximation.

It is this restricted class of functions $\emptyset_{-1}$ and $\emptyset_1$ with which the remainder of our discussion will be concerned.

The Fourier method will be streamlined into a Chebyshev method by adopting approximations to v (x,t) of the form

$$\sum_{i=0}^{2n+1} \sum_{j=1}^{2n} C_{ij}\, T_i\,(x)\, e^{m_j t}. \tag{6}$$

These approximations will be substantially partial sums of an infinite expansion

$$\sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \tilde{C}_{ij}\, T_i\,(x)\, e^{\tilde{m}_j t}. \tag{7}$$

At first sight the number of coefficients involved in (6) appears to be of order n^2, but in fact it is found that C_{ij} may be generated by simple recurrence formulae from just 2n basic coefficients. Thus the Chebyshev method will in effect be one-dimensional.

2. The Basic Chebyshev Method

The form of approximation (6) will not be adopted immediately. Instead, undetermined functions of t will be used initially, and it will be shown that these must necessarily be sums of exponentials.

The heat-conduction problem (1) and (2) is replaced by the following discrete problem:

Find the solution of form

$$v(x,t) = \sum_{i=0}^{n} x^{2i} \left[a_{n-i}(t) + x\, b_{n-i}(t)\right] \qquad \ldots(8)$$

to the partial differential equation

$$v_{xx} - v_t = 2^{1-2n}\, T_{2n}(x) \cdot \left[\tau_1(t) + x\, \tau_2(t)\right], \qquad \ldots(9)$$

subject to the boundary conditions

$$\lambda_{-1} v + \mu_{-1} v_x = \emptyset_{-1}(t) \quad \text{on } x = -1, \qquad \ldots(10a)$$

$$\lambda_1 v + \mu_1 v_x = \emptyset_1(t) \quad \text{on } x = 1, \qquad \ldots(10b)$$

$$v(x,t) = f(x) \text{ at the zeros of } T_{2n}(x) \text{ on } t = 0. \qquad \ldots(10c)$$

The boundary condition (10c) is equivalent, in the case $\emptyset_{-1} \equiv \emptyset_1 \equiv 0$, to the condition

$$v(x,t) = f^*(x) \text{ on } t = 0, \qquad \ldots(10d)$$

where $f^*(x)$ is the polynomial of degree $(2n + 1)$ which interpolates $f(x)$ at the $2n$ zeros of the Chebyshev polynomial $T_{2n}(x)$ subject to the constraints (10a) and (10b) on $t = 0$. The use of the perturbation term on the right of (9) is an obvious generalization of the tau method of Lanczos (1956).

As was inferred in Sec. 1, we shall assume that ϕ_{-1} and ϕ_1 are sums of products of functions like t, e^{-t}, $\cos t$, and $\sin t$, and that $f(x)$ is continuous and of bounded variation on $[-1, 1]$. The restriction on $f(x)$ ensures that it has a uniformly convergent Chebyshev series on $[-1+\epsilon, 1-\epsilon]$ for any $\epsilon > 0$, which in turn gives substance to the process of Chebyshev interpolation used in (10c).

Define the constants u_o, u_1, ..., u_n by the equation

$$2^{1-2n}\, T_{2n}\,(x) \equiv u_o\, x^{2n} + u_1\, x^{2n-2} + \ldots + u_n,$$

so that in particular $u_o = 1$. Then the substitution of (8) into (9) leads, on equating powers of x, to the following system of first order ordinary differential equations.

$$\left.\begin{aligned} - a_0^{(1)} &= \tau_1, \\ (2n-2i+2)(2n-2i+1)\, a_{i-1} - a_i^{(1)} &= \tau_1\, u_i,\ (i=1,\ldots,n) \end{aligned}\right\}\ldots(11)$$

$$\left.\begin{aligned} - b_0^{(1)} &= \tau_2, \\ (2n-2i+3)(2n-2i+2)\, b_{i-1} - b_i^{(1)} &= \tau_2\, u_i,\ (i=1,\ldots,n) \end{aligned}\right\}\ldots(12)$$

Derivatives with respect to t are denoted by any of the notations

$$\frac{d}{dt}\, a_0 \equiv D\, a_0 \equiv a_0^{(1)}, \text{ etc.}$$

The following relations can be verified by substitution into (11) and (12).

$$a_{n-i}^{(n)} = \sum_{j=i}^{n} G_{ij}\, a_0^{(j)}, \qquad i = 0, \ldots, n \qquad \ldots(13)$$

$$b_{n-i}^{(n)} = \sum_{j=i}^{n} H_{ij}\, b_0^{(j)}, \qquad i = 0, \ldots, n \qquad \ldots(14)$$

where $$G_{ij} = \frac{(2n+2i-2j)!}{(2i)!} \; u_{j-i} \qquad \ldots(15)$$

and $$H_{ij} = \frac{(2n+2i-2j+1)!}{(2i+1)!} \; u_{j-i} \qquad \ldots(16)$$

for $j = i, \ldots, n$ and $i = 0, \ldots, n$

If conditions (10a), (10b) are imposed on $v(x, t)$, then the following relations result

$$\left.\begin{aligned} &\sum_{i=0}^{n} (\alpha_i\, a_{n-i} + \beta_i\, b_{n-i}) = \phi_{-1}\, (t) \\ \text{and} \quad &\sum_{i=0}^{n} (\gamma_i\, a_{n-i} + \delta_i\, b_{n-i}) = \phi_1\, (t), \end{aligned}\right\} \qquad \ldots(17)$$

$$\left.\begin{aligned} \text{where } \alpha_i &= \lambda_{-1} - 2i\, \mu_{-1},\quad \beta_i = -\lambda_{-1} + (2i+1)\, \mu_{-1} \\ \gamma_i &= \lambda_1 + 2i\, \mu_1,\quad \delta_i = \lambda_1 + (2i+1)\, \mu_1 \end{aligned}\right\} \qquad \ldots(18)$$

Let P, Q, R, S, E be defined as follows.

$$P(z) \equiv p_0 + p_1\, z + \ldots + p_n\, z^n, \quad Q(z) \equiv q_0 + \ldots + q_n\, z^n,$$

$$R(z) \equiv r_0 + r_1\, z + \ldots + r_n\, z^n, \quad S(z) \equiv s_0 + \ldots + s_n\, z^n,$$

$$E(z) \equiv e_0 + e_1\, z + \ldots + e_{2n}\, z^{2n} \equiv Q(z)\, R(z) - P(z)\, S(z), \qquad \ldots(19)$$

where p_j, q_j, r_j, s_j are some numbers, and z is either a variable or a differential operator. Then, differentiating (17) n times and substituting for $a_{n-i}^{(n)}$ and $b_{n-i}^{(n)}$ from (13) and (14), we obtain the following pair of simultaneous linear ordinary differential equations

$$\left.\begin{aligned} P(D)\, a_0 + Q(D)\, b_0 &= \phi_{-1}^{(n)}\ (t), \\ R(D)\, a_0 + S(D)\, b_0 &= \phi_{1}^{(n)}\ (t), \end{aligned}\right\} \qquad \ldots(20)$$

where D denotes $\frac{d}{dt}$ and

$$p_j = \sum_{i=0}^{j} \alpha_i\, G_{ij}, \qquad q_j = \sum \beta_i\, H_{ij}, \qquad \ldots(21a)$$

$$r_j = \sum \gamma_i\, G_{ij}, \qquad s_j = \sum \delta_i\, H_{ij}. \qquad \ldots(21b)$$

Separating a_0 and b_0 in (20) we obtain

$$E(D)\, a_0 = -S(D)\, \phi_{-1}^{(n)}(t) + Q(D)\, \phi_{1}^{(n)}\ (t), \qquad \ldots(22a)$$

$$E(D)\, b_0 = R(D)\, \phi_{-1}^{(n)}(t) - P(D)\, \phi_{1}^{(n)}\ (t). \qquad \ldots(22b)$$

where E and $e_0, \ldots, e_{2n}$ are defined by (19).

The solutions of (22a) and (22b) may be expressed in the form

$$a_0 = \sum_{j=1}^{2n} k_j\, e^{m_j t} + \xi_0\ (t), \qquad \ldots(23)$$

$$b_0 = \sum_{j=1}^{2n} l_j\, e^{m_j t} + \eta_0\ (t), \qquad \ldots(24)$$

where ξ_0 (t) and η_0 (t) are particular integrals of (22a) and (22b) respectively, k_j and l_j are undetermined constants, and m_j are the 2n roots of the polynomial equation

$$E(z) \equiv e_0 + e_1 z + \ldots + e_{2n} z^{2n} = 0. \qquad \ldots(25)$$

Moreover k_j and l_j may be shown to be connected by the relation

$$k_j\, R(m_j) + l_j\, S\ (m_j) = 0$$

If we denote $R(m_j)$ by R_j and $S(m_j)$ by S_j, then k_j and l_j may be expressed in terms of an unknown h_j by writing

$$k_j = h_j S_j \text{ and } l_j = -h_j R_j. \qquad \ldots(26)$$

The constants m_j are determined by (25), and it remains to determine h_j. But first the remaining coefficients a_1, ..., a_n, b_1, ..., b_n must be expressed in similar forms to (23) and (24), and this can be achieved by applying the equations (11) and (12) to (23) and (24). Suppose that

$$M_j = (m_j)^{-1} \text{ for each } j,$$

then it can be verified that

$$a_i (t) = \sum_{j=1}^{2n} k_j A_{ij} e^{m_j t} + \xi_i (t), \; i = 0, \ldots, n \qquad \ldots(27)$$

$$b_i (t) = \sum_{j=1}^{2n} l_j B_{ij} e^{m_j t} + \eta_i (t), \; i = 0, \ldots, n \qquad \ldots(28)$$

where

$$A_{oj} = 1, \; A_{ij} = (2n-2i+2)(2n-2i+1) M_j A_{i-1,j} + u_i, \qquad \ldots(29)$$

$$B_{oj} = 1, \; B_{ij} = (2n-2i+3)(2n-2i+2) M_j B_{i-1,j} + u_i, \qquad \ldots(30)$$

for $j = 1, 2, \ldots, 2n$ and $i = 1, 2, \ldots, n$,

and $\xi_i (t)$, $\eta_i (t)$ are functions which may be determined explicitly from $\xi_0 (t)$ and $\eta_0 (t)$ by applying the relations (11) and (12).

Moreover, similar formulae can be deduced for τ_1 and τ_2.

$$\tau_1 = - \sum_{j=1}^{2n} k_j m_j e^{m_j t} - \xi_0^{(1)} (t), \qquad \ldots(31)$$

$$\tau_2 = - \sum_{j=1}^{2n} l_j m_j e^{m_j t} - \eta_0^{(1)} (t). \qquad \ldots(32)$$

Assuming that the m_j are all real and negative, we note that τ_1 and τ_2 decay asymptotically to zero if and only if the same is true of $\xi_0^{(1)}$ and $\eta_0^{(1)}$, respectively.

This is true in the trivial case in which ϕ_{-1} and ϕ_1 are identically zero, for the functions $\xi_0, \ldots, \xi_n, \eta_0, \ldots, \eta_n$ are all then identically zero. No formulae will be derived here for ξ_i and η_i in a general case, but in Section 3 below we shall cover in detail the particular case in which ϕ_{-1} and ϕ_1 are both polynomials in t.

Combining (8), (26), (27), and (28) leads to the following formula for v(x,t).

$$v(x,t) \equiv U(x,t) + V(x,t), \qquad \ldots(33)$$

where

$$U(x,t) = \sum_{i=0}^{n} \sum_{j=1}^{2n} x^{2n-2i} h_j e^{m_j t} (S_j A_{ij} - x R_j B_{ij}), \ldots(34)$$

$$\text{and} \quad V(x,t) = \sum_{i=0}^{n} x^{2n-2i} \left[\xi_i (t) + x \eta_i (t)\right] \qquad \ldots(35)$$

The boundary condition (10c) is all that remains to be satisfied in the discrete problem (9), (10a-c). This condition specifies that

$$v(x,o) = f(x) \text{ at the 2n zeros of } T_{2n} (x), \qquad \ldots(36)$$

where v(x,t) is defined by (33). These are just 2n simultaneous linear equations for the determination of the 2n unknowns h_j.

Thus the solution of the discrete problem has reduced to the determination of the quantities m_j and h_j by the solution of the polynomial (25) and the set of linear equations (36).

2.1 Special Cases

If λ_{-1}, λ_1 μ_{-1}, μ_1 satisfy

$$(-\lambda_{-1} + \mu_{-1})\,\lambda_1 - \lambda_{-1}\,(\lambda_1 + \mu_1) = 0, \qquad \ldots(37)$$

then it can be verified that $e_0 = 0$, $e_1 \neq 0$ in (25). Thus (25) has a simple zero root, say m_{2n}. The particular case in which both $(\lambda_1 + \mu_1)$ and $(-\lambda_{-1} + \mu_{-1})$ are zero is of no practical interest. On the other hand, the case $\lambda_{-1} = \lambda_1 = 0$ is of interest and is covered by the special Chebyshev method of Section 4. In all other cases of (37) we simply modify the relations (29) and (30) by setting $A_{i,2n} = B_{i,2n} = 0$ for $i = 0, \ldots, n - 1$.

2.2 Chebyshev Coefficients and Error Estimates

The function $U(x,t)$ of (34) may be rewritten in the form

$$U(x,t) \equiv \sum_{i=0}^{2n+1} \sum_{j=1}^{2n} K_{ij}\, x^i\, e^{m_j t}, \qquad \ldots(38)$$

where K_{ij} are easily computed. The form (38) may in turn be rearranged into the Chebyshev form of (6), namely

$$U(x,t) \equiv \sum_{i=0}^{2n+1} \sum_{j=1}^{2n} C_{ij}\, T_i\,(x)\, e^{m_j t}, \qquad \ldots(39)$$

where C_{ij} are readily generated from K_{ij}.

Similarly $V(x,t)$ may be rewritten as

$$V(x,t) \equiv \sum_{i=0}^{2n+1} \zeta_i\,(t)\, T_i\,(x), \qquad \ldots(40)$$

where $\{\zeta_i\,(t)\}$ are easily calculated from $\{\xi_i\,(t)\}$, $\{\eta_i\,(t)\}$.

Now, from (39), we may write

$$U(x,t) = \sum_{i=0}^{2n+1} D_i(t)\, T_i(x),$$

where $D_i(t) = \sum_{j=1}^{2n} C_{ij}\, e^{m_j t}$

Since $|T_i(x)| \leq 1$, estimates may be made of the discrepancy between the solutions of the discrete problem and the original problem on any line in the t-direction by comparing the coefficients $D_i(t)$ and $\zeta_i(t)$ which are obtained for different choices of n.

Errors could also be estimated by analyzing the perturbation terms on the right hand side of (9), using the formulae (31) and (32). But it seems simpler to compare Chebyshev coefficients.

3. Separation of Steady State and Transient Solutions

The solution v(x,t) of the original problem (1) and (2) of Section 1 may be separated into two distinct parts

$$v(x,t) \equiv U(x,t) + V(x,t) \tag{41}$$

The function V(x,t) is a particular solution of

$$V_{xx} - V_t = o \tag{42}$$

such that
$$\left.\begin{aligned} &\lambda_{-1} V + \mu_{-1} V_x = \emptyset_{-1}(t) && \text{on } x = -1, \\ \text{and}\quad &\lambda_1 V + \mu_1 V_x = \emptyset_1(t) && \text{on } x = 1 \end{aligned}\right\} \tag{43}$$

The function U(x,t) is then determined from

$$U_{xx} - U_t = o \tag{44}$$

subject to
$$\left.\begin{aligned} &\lambda_{-1} U + \mu_{-1} U_x = 0 && \text{on } x = -1, \\ &\lambda_1 U + \mu_1 U_x = 0 && \text{on } x = 1, \\ &U = g(x) && \text{on } t = o, \end{aligned}\right\} \tag{45}$$

where
$$g(x) \equiv f(x) - V(x,o) \tag{46}$$

The term U(x,t) decays exponentially with t and is called the transient solution, while V(x,t) is independent of f(x) and is called the steady state solution. We note that the solution (33) of the discrete problem in section 2 was expressed in an analogous way.

If a steady state solution V(x,t) can be found, then the problem (1) and (2) reduces to a problem (44) and (45) of the same form, but with $\emptyset_{-1}$ and $\emptyset_1$ identically zero. We shall obtain formulae for V(x,t) when $\emptyset_{-1}$ and $\emptyset_1$ are polynomials in t.

For simplicity we solve (42) with the boundary conditions

$$\left.\begin{aligned} &\lambda_o V + \mu_o V_x = \emptyset_o(t) && \text{on } x = o, \\ &\lambda_1 V + \mu_1 V_x = \emptyset_1(t) && \text{on } x = 1 \ . \end{aligned}\right\} \tag{47}$$

The conditions (43) can be converted into (47) by suitable transformations of x and t (but note that λ_1, μ_1, and ϕ_1 are not the same in (43) and (47)).

Suppose that α_k, β_k are not both zero and that

$$\left.\begin{aligned}\phi_o(t) &\equiv \alpha_o + \alpha_1 t + \ldots + \alpha_k t^k, \\ \phi_1(t) &\equiv \beta_o + \beta_1 t + \ldots + \beta_k t^k .\end{aligned}\right\} \qquad (48)$$

Then the solution of (42) and (47) has the form

$$V(x,t) \equiv \sum_{i=o}^{k} x^{2i} \left[y_i(t) + x\, z_i(t) \right] \qquad (49)$$

where $y_i(t) = \sum_{j=o}^{k} Y_{ij}\, t^j$ and $z_i(t) = \sum_{j=o}^{k} Z_{ij}\, t^j$

Algorithms for determining Y_{ij} and Z_{ij} are given below for all relevant cases. Note that, for $k \leq n$, the functions $y_{n-i}(t)$ and $z_{n-i}(t)$ are precisely the functions $\xi_i(t)$ and $\eta_i(t)$ of section 2. (after the relevant transformations)

Constants p_n are defined as follows:

$p_o = 1$; $p_n = (n!)^{-1}$, $n \geq 1$; $p_n = o$, $n < o$.

3.1 $\mu_o \neq o$, λ_o and λ_1 not both zero

Define $\delta_j = (\alpha_j - \lambda_o \gamma_j)/\mu_o$ and

$$Y_{oj} = \gamma_j, \quad Z_{oj} = \delta_j \qquad j = o, \ldots, k \qquad (50)$$

$$\left.\begin{aligned} 2i(2i-1)\, Y_{ij} &= (j+1)\, Y_{i-1,j+1}, \\ (2i+1)2i\, Z_{ij} &= (j+1)\, Z_{i-1,j+1} \end{aligned}\right\} \quad \begin{aligned} j &= o, 1, \ldots, k-i \\ i &= 1, \ldots, k \end{aligned} \qquad (51)$$

γ_j are determined successively from the relations

$$\beta_j = (\mu_o)^{-1} \sum_{i=j}^{k} \frac{i!}{j!} (\gamma_i F_{ij} + \alpha_i G_{ij}), \quad j = k, k-1, \ldots, o \qquad (52)$$

where $F_{ij} = \mu_o \mu_1 p_{r-1} + (\lambda_1 \mu_o - \lambda_o \mu_1) p_r - \lambda_o \lambda_1 p_{r+1}$,

and $G_{ij} = \mu_1 p_r + \lambda_1 p_{r+1}$, for $r = 2i - 2j$

3.2 $\mu_o = o$

Set $\gamma_j = \alpha_j/\lambda_o$, define Y_{ij} and Z_{ij} by (50) and (51), determining δ_j successively from

$$\beta_j = (\lambda_o)^{-1} \sum_{i=j}^{k} \frac{i\,!}{j\,!} (\delta_i F_{ij} + \alpha_i G_{ij}), \quad j = k, k-1, \ldots, o$$

where $F_{ij} = \lambda_o \mu_1 p_r + \lambda_o \lambda_1 p_{r+1}$,

and $G_{ij} = \mu_1 p_{r-1} + \lambda_1 p_r$, for $r = 2i - 2j$.

3.3 $\lambda_o = \lambda_1 = o$

In this case there is an arbitrary constant term in $V(x,t)$, which may be taken as zero.

Set $\delta_j = \alpha_j/\mu_o$, and define Y_{ij} and Z_{ij} by (50) and (51), but with

$$Y_{oo} = o, \; Y_{oj} = 2\,\gamma_{j-1}/j, \; Y_{1j} = \gamma_j \quad . \quad j = o, \ldots, k$$

Determine γ_j from (52) for $j = k, k-1, \ldots, o$ with $F_{ko} = o$ and otherwise

$$F_{ij} = 2\,\mu_o\,\mu_1\,p_{r+1}, \quad G_{ij} = \mu_1\,p_r \quad \text{for } r = 2i - 2j \quad .$$

4. A Special Chebyshev Method for the Radiation Boundary Problem

The analysis of Sec. 2 can be simplified in two significant respects if λ_{-1}, λ_1, μ_{-1}, μ_1 satisfy

$$\lambda_{-1} = \lambda_1, \text{ and } \mu_{-1} = -\mu_1 \quad . \qquad \ldots(53)$$

For the solution of a polynomial (25) of degree 2n then reduces to the solution of two polynomials each of degree n, and the solution of the system (36) of linear equations can be obtained by the use of recurrence relations in order n^2 operations.

Three situations in which (53) apply are as follows.

(a) Radiation at Boundary (compare Example (ii) of Sec. 1)

$$\lambda_{-1} = \lambda_1 = h \neq 0, \; -\mu_{-1} = \mu_1 = 1 \quad . \qquad \ldots(54a)$$

(b) Specified Temperature on Boundary

$$\mu_{-1} = \mu_1 = 0 \qquad \ldots(54b)$$

(c) Specified Flux across Boundary

$$\lambda_{-1} = \lambda_1 = 0 \quad . \qquad \ldots(54c)$$

The problems (b) and (c) are in a sense special cases of (a) in which $h = \infty$, 0 respectively.

When (53) holds in equation (20),

$$P(D) \equiv R(D), \quad Q(D) \equiv -S(D)$$

and hence $2\, P(D)\, a_0 = \phi_{-1}^{(n)} + \phi_1^{(n)}$,

$$2\, Q(D)\, b_0 = \phi_{-1}^{(n)} - \phi_1^{(n)} \; .$$

Thus (27) and (28) reduce to

$$a_i \, (t) = \sum_{j=1}^{n} k_j \, A_{ij} \, e^{m_j t} + \xi_i \, (t), \qquad \ldots(55)$$

and $$b_i(t) = \sum_{j=1}^{n} l_j B_{ij} e^{h_j t} + \eta_i(t), \qquad \ldots(56)$$

where $A_{oj} = 1$, $A_{ij} = (2n-2i+2)(2n-2i+1) M_j A_{i-1,j} + u_i$, ...(57)

$B_{oj} = 1$, $B_{ij} = (2n-2i+3)(2n-2i+2) H_j B_{i-1,j} + u_i$, ...(58)

$M_j = (m_j)^{-1}$ and $H_j = (h_j)^{-1}$

and $\{m_j\}$ and $\{h_j\}$ are the roots of

$P(z) = 0$ and $Q(z) = 0$, with coefficients p_j and q_j given by (21a).

Thus $\{M_j\}$ and $\{H_j\}$ are the roots of

$$p_0 z^n + p_1 z^{n-1} + \ldots + p_n = 0, \qquad \ldots(59)$$

and $$q_0 z^n + q_1 z^{n-1} + \ldots + q_n = 0, \qquad \ldots(60)$$

respectively.

Now suppose that $g(x) \equiv f(x) - V(x,o)$, where $V(x,t)$ is given by (35). Then the set of linear equations (36) reduces to

$$U(x,o) = g(x) \text{ at } x = \pm x_i,\ i = 1, \ldots, n \qquad \ldots(61)$$

where $U(x,t) \equiv \sum_{i=0}^{n} \sum_{j=1}^{n} x^{2n-2i} \left(k_j A_{ij} e^{m_j t} + x l_j B_{ij} e^{h_j t} \right)$

and $\{\pm x_i\}$ are the zeros of $T_{2n}(x)$. To solve the system (61), we first form the unique polynomial

$$F_n(x) \equiv \sum_{i=0}^{n-1} x^{2i} \left(\bar{y}_{n-i} + x \bar{\bar{z}}_{n-i} \right) \qquad \ldots(62a)$$

of degree (2n-1) which interpolates g(x) at the relevant points. The Lagrange formula for (62a) may be expressed as

$$F_n(x) \equiv \sum_{i=1}^{n} \left(G_{i1} + \frac{x}{x_i} G_{i2} \right) L_i^*(x), \qquad \ldots(62b)$$

where $G_{i1} = \frac{1}{2}\left[g(x_i) + g(-x_i)\right]$, $G_{i2} = \frac{1}{2}\left[g(x_i) - g(-x_i)\right]$,

and $L_i^*(x) = \prod_{j \neq i} \left[(x^2 - x_j^2)/(x_i^2 - x_j^2)\right]$.

If a number of problems are to be solved, then the coefficients in $L_i^*(x)$ may be regarded as known, and the computations of $\{\bar{y}_{n-i}\}$ and $\{\bar{z}_{n-i}\}$ from (62a) and (62b) thus involve about $2n^2$ operations each.

Let $\sum_{i=0}^{n} x^{2i}(y_{n-i} + x\, z_{n-i})$ be the polynomial which interpolates $g(x)$ at $\{\pm x_i\}$ subject to the constraints (10a) and (10b) on $t = 0$, with $\emptyset_{-1} \equiv \emptyset_1 \equiv 0$. Then, since

$$2^{1-2n}\, T_{2n}(x) \equiv u_0\, x^{2n} + \ldots + u_n,$$

numbers ν_1 and ν_2 can be found such that

$$\sum x^{2i}(y_{n-i} + x\, z_{n-i}) = \sum x^{2i}(\bar{y}_{n-i} + x\, \bar{z}_{n-i}) + (\nu_1 + \nu_2 x) \sum x^{2i} u_{n-i},$$

where we suppose that $\bar{y}_0 = \bar{z}_0 = 0$. Hence

$$y_i = \bar{y}_i + \nu_1 u_i, \quad z_i = \bar{z}_i + \nu_2 u_i \quad (i = 0, \ldots, n) \qquad \ldots(63)$$

The constraints (10a) and (10b) require that

$$\sum \alpha_i\, y_{n-i} = 0 \text{ and } \sum \beta_i\, z_{n-i} = 0,$$

where α_i and β_i are given by (18). From (63), these equations lead to formulae for ν_1 and ν_2.

$$\nu_1 = \frac{-\sum \alpha_i\, \bar{y}_{n-i}}{\sum \alpha_i\, u_{n-i}}, \qquad \nu_2 = \frac{-\sum \beta_i\, \bar{z}_{n-i}}{\sum \beta_i\, u_{n-i}}. \qquad \ldots(64)$$

Thus y_i and z_i are now explicitly determined by (63), and the equations (61) may be reduced by the formulation (10d) to

$$\left.\begin{aligned} \sum_{j=1}^{n} k_j A_{ij} &= y_i, \\ \sum_{j=1}^{n} l_j B_{ij} &= z_i. \end{aligned}\right\} \quad i = 0, \ldots, n-1 \qquad \ldots(65)$$

Next define r_i and s_i by the following algorithms.

$$r_o = y_o, \quad (2n)!\, r_i = (2n-2i)!\, y_i - \sum_{j=1}^{i} u_j (2n-2j)!\, r_{i-j}\,,$$

$$i = 1, \ldots, n \qquad \ldots(66)$$

$$s_o = z_o, \quad (2n+1)!\, s_i = (2n-2i+1)!\, z_i - \sum_{j=1}^{i} u_j (2n-2j+1)!\, s_{i-j}\,.$$

$$i = 1, \ldots, n \qquad \ldots(67)$$

By making use of the recurrence relations (57) and (58), we can manipulate equations (65) into

$$\left.\begin{aligned} \sum k_j (M_j)^i &= r_i, \\ \sum l_j (H_j)^i &= s_i. \end{aligned}\right\} \quad i = 0, \ldots, n-1 \qquad \ldots(68)$$

Now suppose that $\{M_j\}$ were found from equation (59) in the order $M_n, M_{n-1}, \ldots, M_1$. Then M_j is a root of a deflated polynomial

$$P_j(z) \equiv p_{oj} z^j + p_{1j} z^{j-1} + \ldots + p_{jj} = 0, \qquad \ldots(69)$$

whose coefficients p_{ij} have already been calculated. Since $M_1, \ldots, M_j$ are all roots of (69), we may deduce the following algorithm for the calculation of $\{k_j\}$ from (68).

$$\left.\begin{aligned} k_{j+1} &= \sum_{i=0}^{j} p_{ij}\, r_{j-i} \,/\, P_j\,(M_{j+1}), \\ r_i &= r_i - k_{j+1}\cdot(M_{j+1})^i, \quad i = 0, \ldots, j-1 \end{aligned}\right\} j = n-1, \ldots, 1 \qquad \ldots(70)$$

and $k_1 = r_0$.

A corresponding algorithm defines $\{l_j\}$.

All the computations above, including the solution of polynomials, involve order n^2 operations. Thus this special Chebyshev method has solved the discrete problem (9) and (10a-c) in order n^2 operations.

The case $\lambda_{-1} = \lambda_1 = 0$ (compare section 2.1) requires some modifications. Choose $m_n = 0$, set $k_n = z_n$, $A_{in} = 0$ for all i, and solve (68) for $k_1, \ldots, k_{n-1}$.

5. Summary of the Chebyshev Computations

We summarize the essential parts of the Chebyshev method and count up the number of operations of addition, multiplication, and division required in the derivation and subsequent use of the discrete approximation U(x,t) to the transient solution. Quantities independent of the general problem, such as u_i, $(2i)!$, $L_i^*(x)$, and G_{ij}, will not be included in the count.

The solution of a polynomial was performed by successive deflations, calculating the dominant root at each stage by Bernoulli's method to 2 figure accuracy followed by Newton's method to final accuracy. Let B_i and N_i denote the numbers of iterations which are found necessary in the respective methods for the deflated polynomial of degree i, and define

$$K_j = \sum_{i=1}^{j} \left[(2i+2)\, B_i + (4i+1)\, N_i \right] \qquad \ldots(71)$$

In general it was found that, for $n \leq 10$, the values $B_i = 8$ and $N_i = 3$ were not exceeded.

5.1 Special Chebyshev Method

For the radiation boundary problem in Section 4, the even and odd parts of U (x,t) were produced separately from similar formulae. In Table 5 we analyze the computations required for the unknowns k_j, m_j, A_{ij} involved in the even part of U(x,t). Operations of order n^0 have been ignored. Addition shows that the even (or odd) part of U(x,t) has been derived in

$$11n^2 + 8n + K_n \text{ operations,} \qquad \ldots(72)$$

which for $B_i = 8$ and $N_i = 3$ gives $25n^2 + 41n$ operations.

The evaluation of the even part at p values of x for m values of t involves

$$m[2n^2 + 15n - 1 + 2pn] \text{ operations.} \qquad \ldots(73)$$

We have allowed 15 operations for the evaluation of $k_j e^{m_j t}$.

Table 6

Derivation of Even Solution of Radiation Boundary Problem

Quantities		Relevant Equations	No. of Operations
(i)	α_i	(18)	n
(ii)	p_j	(21a)	n^2+2n
(iii)	M_j, m_j	(59), (69)	$K_n + n^2$
(iv)	$\bar{y}_i$	(62a-b)	$2n^2-n$
(v)	y_i	(63), (64)	$6n$
(vi)	r_i	(66)	n^2+2n
(vii)	k_j	(70)	$3n^2-2n$
(viii)	A_{ij}	(57)	$3n^2$

5.2 General Chebyshev Method

For the general method of Section 2 we do not give a detailed analysis. Even and odd parts are not separable in this case, and the derivation of U(x,t) can be shown to involve

$$18\tfrac{2}{3}\, n^3 + 44\, n^2 + 18n + K_{2n} \text{ operations,} \qquad \ldots(74)$$

which for $B_i = 8$ and $N_i = 3$ becomes

$$18\tfrac{2}{3}\, n^3 + 10\, n^2 + 84n.$$

The evaluation of $U(x,t)$ at p values of x for m values of t involves

$$m\left[8n^2 + 32n - 1 + p(4n+2)\right] \text{ operations.} \qquad \ldots(75)$$

6. Advantages of the Chebyshev Method over the Fourier Method

The following simplifications have arisen from the Chebyshev method:

(i) The Fourier series involves sets of cosines and sines which cannot in general (e.g. (54a)) be related by recurrences. In the Chebyshev method the trigonometric functions have been replaced by polynomials related by recurrences.

(ii) The calculation of a set of integrals has been replaced by a process of polynomial interpolation. For example, a set of Gaussian quadrature formulae are replaced by a single interpolation formula.

(iii) The solution of equations like $(\alpha^2 - h^2) \tan 2\alpha = 2\alpha h$ has been replaced by the solution of polynomials. Trigonometric functions are often computed from a polynomial approximation in any case.

In addition to these three simplifications are the following advantages:

(iv) The Chebyshev series can be expected to converge more rapidly than the Fourier series, especially when $f(x)$ is not truly periodic.

(v) The Chebyshev method could, in principle, be applied for any $\lambda_{-1}, \mu_{-1}, \lambda_1, \mu_1$, whereas the Fourier method has only been applied to mixed boundary problems for which $\lambda_1 \mu_{-1} + \mu_1 \lambda_{-1} = 0$.

6.1 Numerical Example

To compare the convergence rates of the two methods, consider a problem for which the Fourier method is particularly simple.

$$v_{xx} - v_t = 0$$

with $v_x = 0$ on $x = 0$, $v = t^2$ on $x = 1$,

and $v = \cos \frac{1}{2} \pi x$ on $t = 0$

The steady state solution, by Section 3, has the form

$$V(x,t) \equiv \frac{1}{12} (x^2-1)(x^2-5) + t(x^2-1) + t^2$$

and the transient solution from the Fourier method truncates into

$$U(x,t) = e^{-\beta_1 t} \cos \alpha_1 x + 4 \sum_{j=1}^{n} (-1)^j (\alpha_j)^{-5} e^{-\beta_j t} \cos \alpha_j x,$$

where $\alpha_j = (2j-1)\frac{\pi}{2}$, $\beta_j = (\alpha_j)^2$.

By symmetry about x=0, this problem is of form (54b), and the Chebyshev method of Section 4 is applicable. Since g(x) is even, the solution U(x,t) of the relevant discrete problem has form

$$U(x,t) = \sum_{i=0}^{n} \sum_{j=1}^{n} C_{ij} e^{m_j t} T_{2i} (x),$$

dependent on just 2n unknowns $m_1,\ldots,m_n,k_1,\ldots,k_n$.

In tables 1 and 2, the coefficients $-m_j$ and β_j are compared for various values of n, and the convergence of some individual Chebyshev coefficients C_{ij} with n is shown. In Table 3, the maximum errors in U(x,t) are compared for the two methods for various values of n, thus demonstrating the superior convergence of the Chebyshev series.

Table 1 Exponents

	Chebyshev n=3	Chebyshev n=4	Chebyshev n=5		Fourier n=5
$-m_1$	2.467	2.467	2.467	β_1	2.467
$-m_2$	23.711	22.233	22.207	β_2	22.207
$-m_3$	393.822	75.646	62.595	β_3	61.685
$-m_4$		1243.652	178.534	β_4	120.903
$-m_5$			3034.181	β_5	199.859

Table 2

Some Chebyshev Coefficients

	n=3	n=4	n=5
c_{01}	0.2745776	0.2745755	0.2745755
c_{02}	-0.0004911	-0.0004599	-0.0004576
c_{03}	-0.0000019	-0.0000308	-0.0000288
c_{04}		-0.0000003	-0.0000045
c_{05}			-0.0000000

Table 3

Maximum Errors in U(x,t)

n	Chebyshev Method	Fourier Method
3	0.0001	0.00004
4	0.000007	0.00001
5	0.0000016	0.000005
6		0.0000025
7		0.0000013

7. Comparison of Chebyshev and Crank-Nicolson Methods

A comparison is made of results obtained for three typical problems of the radiation boundary type in the case $\phi_{-1} \equiv \phi_1 \equiv 0$.

The approximate number of operations required by the Crank-Nicolson method to calculate the numerical solution U(x,t) at n equally spaced values of x on each of m equally spaced t-lines is

$$3n + m\,(9n - 6),$$

using efficient recurrence relations described in Smith (1965). It is assumed that grid points are excluded at which the solution may be determined by symmetry or from the boundary data. Since m is generally regarded as being of order n^2, the method involves order n^3 operations, compared with order n^2 by (72) for the corresponding Chebyshev method.

A summary of the results of the three examples is given in Table 4. The Crank-Nicolson solutions were generated at intervals of .1 or .2 in x at the following basic t-values:

$$t = .1,\ .2,\ .3,\ .4,\ .5,\ .7,\ 1,\ 1.5,\ 2,\ 3,\ 4.$$

Intermediate steps in t were halved until further divisions produced solutions different by less than 0.0001. An analogous estimate was used in the Chebyshev method, comparing values of U(x,t) obtained at intervals of .2 in x on the line t = .1. Thus both methods were accurate to about 4 decimals for all $t \geq .1$, corresponding to the region in which discontinuities had become smoothed out.

Example A. v_x continuous

Consider again the example of Section 6.1. In this case the accuracy 0.0001 was **obtained** for all t.

Example B. v_x discontinuous

Consider the radiation problem (54a) with h = 1 and initial distribution

$$f(x) \equiv \cos \tfrac{1}{2}\pi\, x + \sin \tfrac{1}{2}\pi\, x.$$

The exponents m_j and h_j of the Chebyshev method, which are independent of f(x), were observed to be converging rapidly to the corresponding exponents $-\alpha_j^2$ of the Fourier series, tabulated in Appendix IV of Carslaw and Jaeger (1959). At the point t = 0, the Chebyshev solution of Table 4 attained its maximum error of .05.

Example C. v discontinuous

Consider problem (54b) with

$$v = 0 \text{ on } x = -1 \text{ and on } x = 1,$$

$$v = 1 \text{ on } t = 0.$$

On t = 0 the Chebyshev solution U(x,o) is that polynomial which interpolates $f(x) \equiv 1$ at the zeros of $T_{2n}(x)$, but which equals zero at x = -1 and 1.

Thus $U(x,o) \equiv 1 - T_{2n}(x)$, and the error on t = 0 has maximum value 1 for every n. However, our results suggest that, like the Fourier series, the Chebyshev solutions converge uniformly with n on the range $[-1, 1]$ of x for every t > 0.

Table 4

Comparison of Methods for Accuracy .0001.

	Chebyshev			Crank-Nicolson		
	n	Number of Operations: Derivation	Number of Operations: Error Estimation	n	Number of t steps	Number of Operations
Example A	3	220	100	10	35	3270
Example B	4	850	360	11	96	9300
Example C	4	430	140	5	58	2430

Thus in these examples the Chebyshev method has proved to be from 4 to 10 times as efficient as Crank-Nicolson, allowing only for a crude error estimate.

Acknowledgements

This research work and the computer time required in it were supported by NASA Grant NsG-398 to the Computer Science Center of the University of Maryland.

References

Carslaw, H. S., and Jaeger, J. C. (1959) Conduction of Heat in Solids, 2nd Edition, Chapter 3, Clarendon Press, Oxford.

Crank, J., and Nicolson, P. (1947) A practical method for numerical evaluation of solutions of partial differential equations of the heat conduction type. Proc. Camb. Phil. Soc. 43, 50-67.

Elliott, D. (1961) A method for numerical integration of the one-dimensional heat equation using Chebyshev series. Proc. Camb. Phil. Soc. 57, 823-832.

Lanczos, C. (1956) Applied analysis. Prentice Hall.

Mason, J. C. (1967) Chebyshev polynomial approximations for the L-membrane eigenvalue problem. S.I.A.M. Journal on Appl. Math. 15, 172-186.

Smith, G. D. (1965) Numerical Solution of Partial Differential Equations. Oxford University Press.